

An Early Warning System for Overcrowding in the Emergency Department

Nathan Hoot, MS¹, Dominik Aronsky, MD, PhD^{1,2}

¹Department of Biomedical Informatics, ²Department of Emergency Medicine
Vanderbilt University Medical Center, Nashville, TN

Overcrowding of emergency departments impedes health care access and quality nationwide. A real-time early warning system for overcrowding may allow administrators to alleviate the problem before reaching a crisis state. Two original probabilistic models – a logistic regression and a recurrent neural network – were created to predict overcrowding crises one hour in the future. The two original and two pre-existing models were validated at 8,496 observation points from January 1, 2006 to February 28, 2006. All models showed high discriminatory ability in terms of area under the receiver operating characteristic curve (logistic regression = .954; recurrent neural network = .957; EDWIN = .879; NEDOCS = .924). At comparable rates of false alarms, the logistic regression gave more advance notice of crises than other models (logistic regression = 62 min; recurrent neural network = 13 min; EDWIN = 0 min; NEDOCS = 0 min). These results demonstrate the feasibility of using models based on key operational variables to anticipate overcrowding crises in real time.

INTRODUCTION

Emergency department (ED) overcrowding is recognized as a national problem that hinders the delivery of emergency medical services [1]. Overcrowding in the ED has been linked to decreased quality of care [2,3], increased costs [4], and patient dissatisfaction [5]. If ED administrators and staff were alerted prior to severe overcrowding, they might be able to intervene before health care quality and access were compromised [6].

There have been no previous attempts, by the knowledge of the authors, to develop an early warning system for ED overcrowding. The goal of the present study is to evaluate the feasibility of implementing models of ED overcrowding – including a logistic regression and a recurrent neural network – as part of a real-time early warning system.

BACKGROUND

ED Overcrowding

A conceptual model has been proposed to differentiate between input, throughput, and output components of ED overcrowding [7]. The input component describes aspects of patient demand that cause overcrowding. The throughput component considers bottlenecks in ED operations that lead to overcrowd-

ing. The output component includes adverse consequences of overcrowding. A panel of experts identified by consensus 38 operational variables that reflect the input, throughput, and output components [8].

In an effort to quantify ED crowding, two measures have been proposed: The Emergency Department Work Index (EDWIN) is a conceptual model, which was defined using expert opinion [9]. The EDWIN score has been shown to correlate with impressions of crowding by physicians and nurses. The National Emergency Department Overcrowding Scale (NEDOCS) is a linear regression model that associates five operational variables with the degree of crowding assessed by physicians and nurses [10].

Early Warning System

An early warning system must incorporate two components: 1) a clear definition of a crisis period, and 2) a means of predicting crises [11]. For the present study, a crisis period of ED overcrowding will be defined as a period when ambulances are diverted to nearby hospitals. Policy at the authors' institution allows for the ED to go on diversion if any of the following criteria apply, and are not expected to be remedied within one hour: 1) all critical care beds in the ED are occupied, patients are occupying hallway spaces, and at least 10 patients are waiting, 2) an acuity level exists that places additional patients at risk, or 3) all monitored beds within the ED are full.

Both EDWIN and NEDOCS could serve as means of predicting overcrowding crises, as the authors of both papers acknowledged [9,10]. One limitation of these models is their use of a subjective dependent variable: clinician assessments of overcrowding. Due to this, the present study will detail the development of two original models using the more objective dependent variable of ambulance diversion.

Evaluation of an early warning system, like diagnostic and prognostic systems, considers the measures of sensitivity and specificity. Beyond this, the evaluation must address the question, "How far in advance can the system anticipate a crisis?" This measure, called timeliness, is defined as the time lapse between when a system first detects a coming crisis, and when the crisis begins [12]. The optimal operating point for an early warning system must be chosen based on the tradeoff between timeliness and the rate of false alarms.

Design Objectives

The system was designed with a central focus: to be practical, it must operate in real time. This goal carries significant information needs. It requires the availability of real-time data feeds to detect crises, as well as a method of instantly disseminating warnings when crises are imminent. Other researchers have stressed the importance of a real-time information technology (IT) infrastructure for other early detection applications, such as monitoring of bioterrorism [13]. At the authors' institution, much of the necessary infrastructure is already in place. The proposed early warning system takes advantage of this.

METHODS

Study Setting and Population

Vanderbilt University Medical Center is a tertiary care academic medical center in an urban setting. The medical center includes both adult and pediatric ED's, and all subsequent discussion in this paper will pertain specifically to the adult ED.

The hospital IT infrastructure includes an electronic health record and computerized physician order entry. An electronic whiteboard display provides an instant summary of ED operational status, helping staff to manage patients and workflow. The whiteboard obtains information from a database that is updated multiple times each minute. The applications are all integrated, enabling real-time collection and analysis of ED operational data.

The study period was defined as the six-month period lasting from September 1, 2005 to February 28, 2006. The first four months will be referred to as the training period, and the last two months will be referred to as the validation period. Operational variables were calculated at every 10-minute interval during the study period. The local Institutional Review Board approved this research.

Data Collection

Operational data were obtained from hospital databases; from staffing schedules of attendings, residents, and nurses; and from diversion logs. The data collected are standard operational data for ED research [8]. The variables used for modeling are shown in table 1.

Variables with potentially ambiguous meanings are defined as follows: Each patient's acuity level is based on the Emergency Severity Index (ESI), a 5-point scale that ranks patients from most urgent (1) to least urgent (5). The mean acuity level is calculated based on all patients in the ED. Occupancy level is defined as the number of patients in the treatment area divided by the number of licensed ED beds. This value may exceed 100% when patients are doubled in rooms or placed in a hallway. Boarding pa-

Table 1. Adult ED operational variables (9/05 - 12/05)

Characteristic	No Diversion n = 14,476	Diversion n = 3,092
Registrations in last hour (#)	5 (3 - 8)	5 (3 - 7)
Discharges in last hour (#)	5 (3 - 7)	6 (4 - 8)
Mean acuity level (ESI)	2.52 \pm 0.15	2.47 \pm 0.13
Occupancy level (%)	66 (51 - 80)	87 (78 - 93)
Average length of stay (h)	4.3 (3.4 - 5.8)	7.1 (5.1 - 9.0)
Waiting patients (#)	1 (0 - 3)	7 (2 - 11)
Average waiting time (min)	4 (0 - 17)	61 (24 - 89)
Boarding patients (#)	5 (3 - 9)	14 (10 - 17)
Average boarding time (h)	2.9 (1.4 - 6.1)	8.4 (5.0 - 11.0)
Attendings on duty (#)	3.1 \pm 0.9	3.4 \pm 0.8
Residents on duty (#)	5.4 \pm 0.5	5.6 \pm 0.5
Nurses on duty (#)	13.6 \pm 1.7	14.3 \pm 1.6
Medical-surgical diversion	5%	24%
Critical care diversion	1%	3%

Observations were made at 10-minute intervals during the training period. Descriptions are presented as percentages for discrete variables, mean \pm SD for normally distributed continuous variables, and median (IQR) for skewed variables.

tients are those for whom hospital admission has been requested, but have not yet been discharged from the ED. The diversion status was recorded for other hospital services that affect ED operations.

Descriptive statistics for all operational variables were calculated for the training period, stratified according to ED diversion status. The mean number of diversion episodes per week, mean duration per episode, and percentage of time on diversion were calculated for the training and validation periods.

Data Preparation

Since some of the operational variables were likely to be collinear, the data were pre-processed by principal component analysis before any statistical models – specifically, the logistic regression and recurrent neural network – were created [14]. By placing a linear restriction on variable correlations, this technique finds a transformation that compresses the maximum amount of variance in a set of data into a few principal components [15]. The principal component analysis was performed using the data from the training period. The amount of variance explained by each component was visualized on a scree plot. The eigenvectors from the top six principal components were stored in a transformation matrix. This matrix was used to transform the raw data from both the training and validation periods into the corresponding principal components. All data processing and analyses were performed with Matlab (version 6.5, <http://www.mathworks.com>).

Statistical Modeling

A logistic regression model was fitted using data from the training period to predict the likelihood that the ED would be on diversion one hour later. The independent variables for each time point were the six principal components, transformed from all 14 operational variables by the method described above. The diversion status of the ED one hour after each observation was the dependent variable.

A recurrent neural network was trained using the same data as the logistic regression. Each neuron in the hidden layer was linked to itself by a recurrent connection, allowing the network to discern temporal structure within data [16]. The hidden layer of the network contained three neurons. The principal components were scaled to the range 0..1, and standard backpropagation was used for training until the mean squared error reached 0.07. To avoid overfitting, ten networks were trained as a committee and their outputs were averaged to obtain final predictions [17].

The number of principal components, the width of the neural network hidden layer, and the duration of neural network training were all tuned in preliminary experiments, which used only data from the training period. All data from the validation period were reserved for the subsequent model validation.

Outcome Measures

The logistic regression and the recurrent neural network were used to predict ED diversion status one hour in the future for every observation in the validation period. Likewise, the values of EDWIN [9] and NEDOCS [10] were calculated for the validation period, according to the published formulas.

Each model's ability to predict future diversion status was analyzed using receiver operating characteristic (ROC) curves [18]. An ROC curve plots (1 - specificity) against sensitivity for all possible thresholds in a binary classification task. The area under an ROC curve (AUC) represents the overall discriminatory ability of a test, where a value of 1.0 denotes perfect ability and a value of 0.5 denotes no ability. Each model's threshold was fixed to achieve 90% sensitivity. At the threshold, the specificity, predictive values, and likelihood ratios were calculated.

The feasibility of each model to be used in an early warning system was analyzed using activity monitoring operating characteristic (AMOC) curves [19]. An AMOC curve plots false alarm rates against timeliness scores for all possible thresholds. Consecutive false alarms were treated as independent events, instead of as a sustained signal. The false alarm rate is normalized per unit time, in this case per hour. The timeliness score may be interpreted here as the median warning time given, within a maximum

three hour window. Although the models were trained using the one-hour time point, their predictions may be useful even further in advance.

RESULTS

There were 15,687 total patient visits to the ED during the training period and 7,726 during the validation period. Recording observations at 10-minute intervals resulted in 17,568 data points for the training period and 8,496 for the validation period.

Descriptive statistics for operational variables during the training period are listed in table 1. The incidence and length of diversion episodes for the training and validation periods are shown in table 2.

Table 2. Ambulance diversion episodes (9/05 - 2/06)

	Training (9/05-12/05)	Validation (1/06-2/06)
Mean episodes per week (#)	2.9	2.6
Mean duration per episode (h)	10.3	16.3
Total time on diversion	18%	25%

Principal component analysis showed that 99% of the total variance among the operational variables could be explained with six principal components. The beta coefficients for the corresponding six independent variables in the logistic regression model were statistically significant ($p < .001$ for each).

The ROC curves for the logistic regression, the recurrent neural network, EDWIN, and NEDOCS are shown in figure 1. The area under the curve was .954 for the logistic regression, .957 for the recurrent neu-

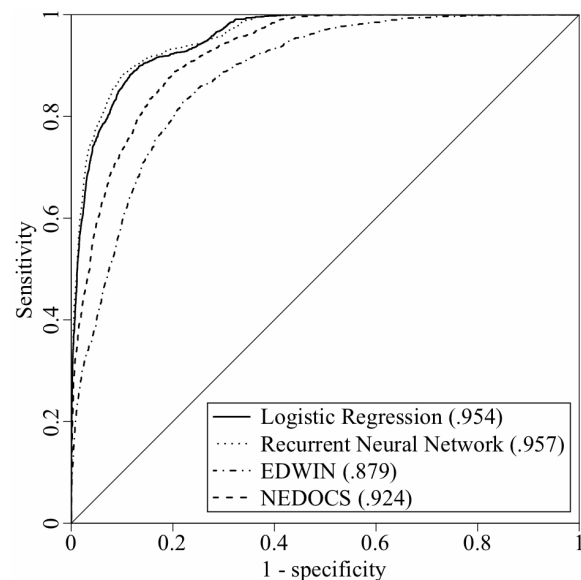


Figure 1. Receiver operating characteristic curves of early warning system predictions. The AUC of each model is shown in parentheses in the lower right.

ral network, .879 for EDWIN, and .924 for NEDOCS. At a fixed sensitivity level of 90%, operating characteristics for each model are shown in table 3.

Table 3. Operating characteristics at fixed 90% sensitivity

	Spec	PPV	NPV	LR+	LR-
Regression	86%	69%	96%	6.59	0.12
Network	87%	71%	96%	7.18	0.11
EDWIN	68%	48%	95%	2.78	0.15
NEDOCS	78%	58%	96%	4.03	0.13

Regression = logistic regression; Network = recurrent neural network; Spec = specificity; PPV = positive predictive value; NPV = negative predictive value; LR+ = positive likelihood ratio; LR- = negative likelihood ratio

The AMOC curves for the four models are presented in figure 2. The logistic regression had the highest timeliness at all rates of false alarms. The recurrent neural network and NEDOCS had lower timeliness and were comparable to each other. EDWIN had the lowest timeliness of the four models.

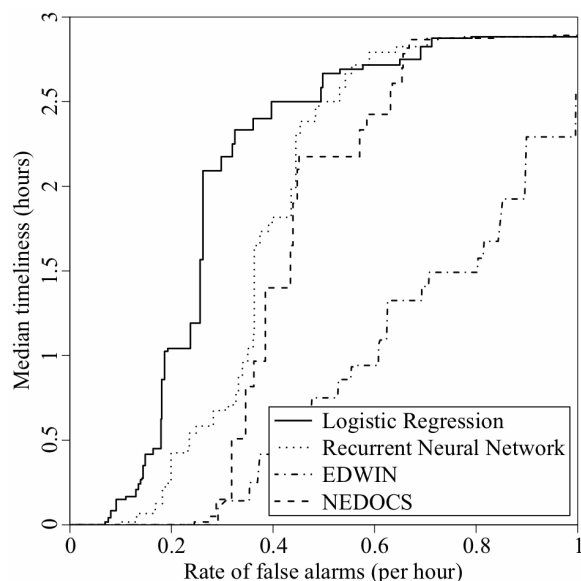


Figure 2. Activity monitoring operating characteristic curves of early warning system predictions. A higher value of timeliness denotes a greater amount of warning time prior to crises.

DISCUSSION

Interpretation of Findings

The results from the validation period show that the logistic regression and the recurrent neural network had very good discriminatory power in predicting diversion status one hour in the future. At a fixed sensitivity level, the logistic regression and recurrent neural network had similar operating characteristics. The logistic regression was more timely in detecting

diversion episodes at all rates of false alarms. Comparison with the pre-existing indices, EDWIN and NEDOCS, reveals that neither of them achieved greater timeliness than the logistic regression, even though both had good discriminatory ability.

It warrants brief mention that the recurrent neural network did not exceed the performance of the logistic regression. Given the assumptions underlying each model, this may question the necessity of considering non-linear relationships, interaction terms, or temporal structure in order to make good predictions with this data set. However, this does not rule out the possibility that other classifiers or neural network architectures may perform better, so model complexity will remain a question for further investigation.

Practical Implications

The results of the validation suggest that the logistic regression may be well-suited for deployment in an early warning system at the local institution. It can provide an hour of advance notice prior to the start of a diversion episode, with a relatively low rate of false alarms. This hour of warning may allow ED administrators and staff to intervene before access and quality of care become compromised. Possible interventions include calling in reserve personnel, opening auxiliary treatment bays, clearing out hospital beds, or deferring care of low acuity patients [6]. Thus, administrators can be proactive, rather than merely reactive, in the face of an overcrowding crisis.

Even with an early warning system in place, some causes of severe overcrowding, such as mass casualties, will likely remain difficult to foresee and remedy, but such events represent a minority of ED crises. Fortunately, based on the results of model validation, most of the diversion episodes at the local institution seem to arise from predictable causes.

Study Limitations

The present study has three limitations that merit discussion. First, because no universal definition of overcrowding exists, ambulance diversion status was used to define crises of overcrowding. Other researchers have identified diversion as the most useful functional definition of overcrowding [20]. Furthermore, the authors' institution has clearly defined criteria for when ambulance diversion is appropriate. Moreover, every decision to initiate diversion undergoes later review, to maintain consistency. Although not all hospitals in the nation allow for ambulance diversion, the operational criteria that define the local policy could be easily applied to other hospitals. On these grounds, the use of ambulance diversion as a surrogate for overcrowding was considered appropriate for the present study.

Second, the serial correlation between consecutive 10-minute observations may affect the independence assumption of the logistic regression model. However, the design of the present study focused on real-world application using independent validation, and a model may still have practical utility when its theoretical assumptions are not fully satisfied.

Third, the logistic regression and recurrent neural network were developed and validated at a single institution. The design of the present study does not allow for comment on how well they would generalize to other institutions. By design, these models were developed using common ED operational variables that can be obtained and processed in real time at any hospital, given adequate IT infrastructure. Further research will be necessary to determine the applicability of the models to other institutions.

Future Directions

To assess the utility of the early warning system, a cost function must be associated with the alarms generated. This will depend on how the institution responds to alarms. It must take into account the costs incurred by calling in reserve staff or opening additional treatment bays, and it must consider the benefit achieved by maintaining efficient operations in the ED. This benefit may take the form of increased revenue for the hospital or intangible societal benefits arising from improved access to emergency care. The intervention policy and cost function must be developed jointly with ED administrators. These elements are necessary to determine the optimal operating point of the system, so that the expected utility will be maximized.

CONCLUSION

Models based on ED operational data may be used as an early warning system for diversion. The logistic regression model predicted diversion with better timeliness and fewer false alarms than other models of overcrowding.

ACKNOWLEDGEMENTS

The first author was supported in part by the National Library of Medicine grant LM07450-02. The second author was supported by the National Library of Medicine R21 grant LM009002-01.

REFERENCES

1. Derlet RW. Overcrowding in the ED. *J Emerg Med*. 1992;10:93-94.
2. Schull MJ, Morrison LJ, Vermeulen M, et al. Emergency department overcrowding and ambulance transport delays for patients with chest pain. *CMAJ*. 2003;168:277-283.
3. Hwang U, Richardson LD, Sonuyi TO, et al. The effect of emergency department crowding on the management of pain in older adults with hip fracture. *J Am Geriatr Soc*. 2006;54:270-275.
4. Bayley MD, Schwartz JS, Shofer FS, et al. The financial burden of emergency department congestion and hospital crowding for chest pain patients awaiting admission. *Ann Emerg Med*. 2005;45:110-117.
5. Jenkins MG, Roche LG, McNicholl BP, et al. Violence and verbal abuse against staff in accident and emergency departments: a survey of consultants in the UK and the Republic of Ireland. *J Accid Emerg Med*. 1998;15:262-265.
6. Forster AJ. An agenda for reducing emergency department crowding. *Ann Emerg Med*. 2005;45: 479-81.
7. Asplin BR, Magid DJ, Rhodes KV, Solberg LI, Lurie N, Camargo CA Jr. A conceptual model of emergency department crowding. *Ann Emerg Med*. 2003;42:173-80.
8. Solberg LI, Asplin BR, Weinick RM, Magid DJ. Emergency department crowding: consensus development of potential measures. *Ann Emerg Med*. 2003;42:824-34.
9. Bernstein SL, Verghese V, Leung W et al. Development and validation of a new index to measure emergency department crowding. *Acad Emerg Med*. 2003;10:938-942.
10. Weiss SJ, Derlet R, Arndahl J et al. Estimating the degree of emergency department overcrowding in academic medical centers. Results from the National ED Crowding Study (NEDOCS). *Acad Emerg Med*. 2004;11:38-50.
11. Edison HJ. Do indicators of financial crises work? An evaluation of an early warning system. *Int J Fin Econ*. 2003;8:11-53.
12. Wagner MM, Tsui FC, Espino JU, et al. The emerging science of very early detection of disease outbreaks. *J Public Health Manag Pract*. 2001;7:51-9.
13. Lober WB, Karras BT, Wagner MM, et al. Roundtable on bioterrorism detection: information system-based surveillance. *J Am Med Inform Assoc* 2002;9:105-15.
14. Harrell FE. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. New York: Springer; 2001.
15. Jolliffe IT. Principal component analysis. New York: Springer; 2002.
16. Elman JL. Finding structure in time. *Cognitive Science*. 1988;14:179-212.
17. Bishop CM. Neural networks for pattern recognition. New York: Oxford University Press; 1995.
18. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143:29-36.
19. Fawcett T, Provost F. Activity monitoring: noticing interesting changes in behavior. In: Madigan C (ed). 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 1999;53-62.
20. Schull MJ, Slaughter PM, Redelmeier DA. Urban emergency department overcrowding: defining the problem and eliminating misconceptions. *Can J Emerg Med*. 2002;4:76-83.